

# Data Stewardship Body of Knowledge



eLearningCurve

## **1. Data Stewardship Fundamentals**

### **1.1. Definitions**

- 1.1.1. Stewardship
- 1.1.2. Data Stewardship

### **1.2. Data Stewardship Organizations**

- 1.2.1. Kinds of Data Stewards
  - 1.2.1.1. Kinds of Business Stewards
  - 1.2.1.2. Kinds of IT Stewards
  - 1.2.1.3. Differences of Business and IT Stewards
- 1.2.2. Roles and Responsibilities
- 1.2.3. Reporting and Relationships

### **1.3. Data Steward Characteristics**

- 1.3.1. Knowledge and Experience
- 1.3.2. Skills and Competencies
  - 1.3.2.1. Data Skills
  - 1.3.2.2. Technical Skills
  - 1.3.2.3. Human Skills
  - 1.3.2.4. Facilitation Skills
  - 1.3.2.5. Communication Skills

## **2. Data Management Processes**

### **2.1. Architectural Processes**

- 2.1.1. Enterprise and Subject Area Modeling
- 2.1.2. Data Mapping and Consolidation
- 2.1.3. Data Flow
- 2.1.4. Data Pipelines

### **2.2. Utilization Processes (CRUD)**

- 2.2.1. Create Processes and Practices
- 2.2.2. Retrieval and Reporting Processes and Practices
- 2.2.3. Update Processes and Practices
- 2.2.4. Delete and Archive Processes and Practices
- 2.2.5. Business Process Management

### **2.3. Custodial Processes**

- 2.3.1. Database Administration
- 2.3.2. Security Administration and Access Authorization
- 2.3.3. Backup, Recovery, and Business Continuity

### **2.4. Data Lifecycle Processes and Enabling Technologies**

- 2.4.1. Data Ingestion
- 2.4.2. Data Cataloging
- 2.4.3. Data Refinement
- 2.4.4. Data Preparation
- 2.4.5. Data Delivery

- 2.4.6. Analysis and Reporting
- 2.4.7. Data Operations (DataOps)
- 2.5. *Data Sharing*
  - 2.5.1. Curated Data and Data Products
  - 2.5.2. Open Data
  - 2.5.3. Data Marketplaces
    - 2.5.3.1. Enterprise (internal) data marketplace
    - 2.5.3.2. B2B data exchange marketplace
    - 2.5.3.3. Commercial data marketplace
  - 2.5.4. APIs and Data Services
- 2.6. *Data Risk Management*
  - 2.6.1. Kinds of Data Risks
  - 2.6.2. Risk Assessment
  - 2.6.3. Risk Prevention
  - 2.6.4. Data Breach Detection and Remediation

### **3. Information Management Concepts**

- 3.1. *Types of Data and Information*
  - 3.1.1. Operational and Analytical
  - 3.1.2. Event and Reference
  - 3.1.3. Structured and Unstructured
  - 3.1.4. Transactional Data
  - 3.1.5. Master Data
- 3.2. *Types of Data Stores*
  - 3.2.1. Application Databases
  - 3.2.2. Departmental and End-User Databases
  - 3.2.3. Data Warehouses and Data Marts
  - 3.2.4. Operational Data Stores
  - 3.2.5. Master Data Hubs
  - 3.2.6. Data Lakes
  - 3.2.7. Analytic Sandboxes
- 3.3. *Types of Databases*
  - 3.3.1. Relational Databases
  - 3.3.2. NoSQL Databases
- 3.4. *Common Uses of Data*
  - 3.4.1. Record Keeping and Audit Trail
  - 3.4.2. Reporting and Information
  - 3.4.3. Measurement and Monitoring
  - 3.4.4. Analysis and Discovery
  - 3.4.5. Machine Learning and Automation
- 3.5. *Business Data Flow*
  - 3.5.1. The Business Value of Data

- 3.5.2. Data Sources and Data Acquisition
- 3.5.3. Data Providers and Consumers
- 3.5.4. Data Flow through Organizations and Systems
- 3.5.5. Data Conversion and Consolidation
- 3.5.6. Data Replication and Redundancy
- 3.5.7. Data Sharing and Interfaces
- 3.5.8. Data Disposal and Destruction

### 3.6. *Information Management Disciplines*

- 3.6.1. Data Modeling
- 3.6.2. Metadata Management
- 3.6.3. Content Management
- 3.6.4. Enterprise Information Management
- 3.6.5. Data Quality
- 3.6.6. Data Governance
- 3.6.7. Data Integration
- 3.6.8. Data Warehousing
- 3.6.9. Master Data Management

### 3.7. *Data Analytics*

- 3.7.1. Business Intelligence
- 3.7.2. Business Analytics
- 3.7.3. Performance Management
- 3.7.4. Data Mining
- 3.7.5. Predictive Analytics
- 3.7.6. Data Science

## 4. **Data Quality**

### 4.1. *Quality Management Basics*

- 4.1.1. Quality Perspectives
  - 4.1.1.1. Expectations
  - 4.1.1.2. Purpose
  - 4.1.1.3. Specifications
  - 4.1.1.4. Defects
- 4.1.2. Quality Management Terminology
  - 4.1.2.1. Quality Control (QC)
  - 4.1.2.2. Quality Assurance (QA)
  - 4.1.2.3. Waste and rework
  - 4.1.2.4. Inspection, Correction, Prevention
- 4.1.3. Quality Management Methodologies
  - 4.1.3.1. Total Quality Management (TQM)
  - 4.1.3.2. Statistical Process Control (SPC)
  - 4.1.3.3. Six Sigma

### 4.2. *Data Quality Concepts and Principles*

- 4.2.1. Data Quality Definitions
  - 4.2.1.1. Defect Free
  - 4.2.1.2. Conform to Specifications
  - 4.2.1.3. Fit to Purpose

- 4.2.1.4. Meet Customer Expectations
- 4.2.2. Common Causes of Data Quality Problems
- 4.2.3. Costs and Benefits of Data Quality

#### 4.3. *Data Quality Dimensions*

- 4.3.1. Content Quality (Correctness)
  - 4.3.1.1. Accuracy
  - 4.3.1.2. Completeness
  - 4.3.1.3. Precision
  - 4.3.1.4. Granularity
  - 4.3.1.5. Consistency
- 4.3.2. Structural Quality (Integrity)
  - 4.3.2.1. Identity
  - 4.3.2.2. Reference
  - 4.3.2.3. Cardinality
  - 4.3.2.4. Dependency
  - 4.3.2.5. Inheritance
  - 4.3.2.6. Domain of Values
- 4.3.3. Temporal Quality (Timeliness)
  - 4.3.3.1. Currency
  - 4.3.3.2. Retention
  - 4.3.3.3. Continuity
  - 4.3.3.4. Precedence
- 4.3.4. Business Quality (Value)
  - 4.3.4.1. Aligned
  - 4.3.4.2. Trusted
  - 4.3.4.3. Understandable
  - 4.3.4.4. Reliable
  - 4.3.4.5. Compliant
  - 4.3.4.6. Useful
- 4.3.5. Usage Quality (Usability)
  - 4.3.5.1. Available
  - 4.3.5.2. Accessible
  - 4.3.5.3. Navigable
  - 4.3.5.4. Recoverable
  - 4.3.5.5. Secure
  - 4.3.5.6. Private
- 4.3.6. Presentation Quality (Communication)
  - 4.3.6.1. Clear
  - 4.3.6.2. Organized
  - 4.3.6.3. Non-Ambiguous
  - 4.3.6.4. Tool-Integrated
  - 4.3.6.5. Media-Fit

#### 4.4. *Data Quality Processes and Projects*

- 4.4.1. Data Quality Assessment
- 4.4.2. Root Cause Analysis
- 4.4.3. Data Cleansing
- 4.4.4. Data Enrichment

- 4.5. *Data Quality in IT Processes and Projects*
  - 4.5.1. Data Quality in Application and Database Projects
  - 4.5.2. Data Quality in Data Conversion and Consolidation Projects
  - 4.5.3. Data Quality in Data Sharing and Data Interfaces
- 4.6. *Data Quality for Big Data*
  - 4.6.1. Big Data Quality Criteria
  - 4.6.2. Big Data Quality Assessment
- 4.7. Measuring Data Quality
  - 4.7.1. Data Quality Metrics
  - 4.7.2. Data Quality Scorecard

## **5. Data Integration**

- 5.1. *Data Integration Processes*
  - 5.1.1. Understanding Data
  - 5.1.2. Capturing Data
  - 5.1.3. Transforming Data
  - 5.1.4. Delivering Data
- 5.2. *Data Integration Methods*
  - 5.2.1. ETL/ELT
  - 5.2.2. Virtualization and Federation
  - 5.2.3. Stream Integration
- 5.3. *Data Freshness*
  - 5.3.1. Real-Time and Near Real-Time
  - 5.3.2. Streaming Data
  - 5.3.3. Change Data Capture
  - 5.3.4. Data Latency
- 5.4. *Ensuring Data Quality in Data Integration*
  - 5.4.1. Data Cleansing
  - 5.4.2. Error and Change Monitoring

## **6. Data Governance**

- 6.1. *Data Governance Basics*
  - 6.1.1. Definitions
  - 6.1.2. Purpose
  - 6.1.3. What to Govern and What not to Govern?
- 6.2. *Components of Data Governance*
  - 6.2.1. Governance Goals
    - 6.2.1.1. Quality
    - 6.2.1.2. Security

- 6.2.1.3. Compliance
  - 6.2.1.4. Standardization
  - 6.2.1.5. Usage
  - 6.2.1.6. Value
  - 6.2.1.7. Business Impact
  - 6.2.1.8. Data Ethics
  - 6.2.2. Governance Elements
    - 6.2.2.1. Roles
    - 6.2.2.2. Responsibilities
    - 6.2.2.3. Decision Authority
    - 6.2.2.4. Accountability
  - 6.2.3. Policies
    - 6.2.3.1. Quality
    - 6.2.3.2. Sensitivity
    - 6.2.3.3. Privacy
    - 6.2.3.4. Security
    - 6.2.3.5. Access
    - 6.2.3.6. Retention
    - 6.2.3.7. Disposal
  - 6.2.4. Standards
    - 6.2.4.1. Naming
    - 6.2.4.2. Definition
    - 6.2.4.3. Architecture
    - 6.2.4.4. Transfer
    - 6.2.4.5. Integration
  - 6.2.5. Technology
    - 6.2.5.1. Data Governance Enabling
    - 6.2.5.2. Data Quality
    - 6.2.5.3. Workflow
    - 6.2.5.4. Collaboration
- 6.3. *Data Governance Programs*
- 6.3.1. Business Drivers for Governance
    - 6.3.1.1. Legal
    - 6.3.1.2. Regulatory
    - 6.3.1.3. Financial
    - 6.3.1.4. Operational
    - 6.3.1.5. Competitive
  - 6.3.2. Data Governance Roles
    - 6.3.2.1. Sponsors
    - 6.3.2.2. Owners
    - 6.3.2.3. Data Stewards
    - 6.3.2.4. Custodians
    - 6.3.2.5. Curators
    - 6.3.2.6. Stakeholders
    - 6.3.2.7. Data Officers
    - 6.3.2.8. Councils, Committees, Competency Centers
- 6.4. *Executing Data Governance*
- 6.4.1. Getting Started
  - 6.4.2. Funding and Sponsorship

- 6.4.3. Day-to-Day Governance
- 6.4.4. Sustaining Governance
- 6.4.5. Monitoring Governance: Measures, Metrics, Maturity
- 6.4.6. Evolving Governance
- 6.4.7. Stakeholder Communications
  
- 6.5. *Data and Capabilities*
  - 6.5.1. Data Dependent Business Capabilities
  - 6.5.2. Technological Data Capabilities
  - 6.5.3. Data Management Capabilities
    - 6.5.3.1. The Data Capabilities Assessment Model (DCAM)
  
- 6.6. *Data Literacy*
  - 6.6.1. Understanding Data
  - 6.6.2. Reading Charts and Graphs
  - 6.6.3. Creating Charts and Graphs
  - 6.6.4. Data Storytelling
  
- 6.7. *Modernizing Data Governance*
  - 6.7.1. The Need for Modernization
    - 6.7.1.1. Big Data and Governance
    - 6.7.1.2. Agile and Governance
    - 6.7.1.3. Self-Service and Governance
  - 6.7.2. The Human Side of Data Governance

## **7. Metadata Management**

- 7.1. *Metadata Concepts*
  - 7.1.1. Metadata Definition
  - 7.1.2. Roles of Metadata
    - 7.1.2.1. Classify
    - 7.1.2.2. Describe
    - 7.1.2.3. Guide
    - 7.1.2.4. Control
  - 7.1.3. Types of Metadata
    - 7.1.3.1. Business Metadata
    - 7.1.3.2. Technical Metadata
  - 7.1.4. Metadata Standards and Practices
  
- 7.2. *Data Modeling*
  - 7.2.1. Data Model Principles
  - 7.2.2. Types of Data Models
    - 7.2.2.1. ER vs. Dimensional
    - 7.2.2.2. Logical vs. Physical
    - 7.2.2.3. Who and Why for each Type of Model
  - 7.2.3. Data Modeling and Big Data
    - 7.2.3.1. Modeling Perspective and Purpose
    - 7.2.3.2. Schema-on-Read vs. Schema-on-Write
    - 7.2.3.3. NoSQL Data Modeling



- 7.2.4. Complementary Models
  - 7.2.4.1. Subject Area Models
  - 7.2.4.2. State Transition Models
- 7.2.5. Reading Data Models
- 7.3. *Data Profiling*
  - 7.3.1. Roles of Profiling
    - 7.3.1.1. Support of Data Quality Management
    - 7.3.1.2. Support of Master Data Management
    - 7.3.1.3. Support of Data Migration
    - 7.3.1.4. Support of Data Integration
  - 7.3.2. Profiling Techniques
    - 7.3.2.1. Column Profiling
    - 7.3.2.2. Profiling Data Models
    - 7.3.2.3. Profiling Time-Dependent Data
    - 7.3.2.4. Subject Profiling
    - 7.3.2.5. Profiling State-Transition Models
    - 7.3.2.6. Attribute Dependency Profiling
    - 7.3.2.7. Dynamic Data Profiling
  - 7.3.3. Profiling Challenges
  - 7.3.4. Profiling for Big Data
- 7.4. Data Curation and Cataloging
  - 7.4.1. Why Data Curation
  - 7.4.2. Kinds of Data Curators
  - 7.4.3. Why Data Cataloging
  - 7.4.4. Metadata and the Data Catalog
- 7.5. *Taxonomies and Ontologies*
  - 7.5.1. Taxonomy Definition and Purpose
  - 7.5.2. Ontology Definition and Purpose
  - 7.5.3. Taxonomy and Ontology Uses in Data Management and Data Science

## **8. Master and Reference Data Management**

- 8.1. *Master Data Management (MDM) Basics*
  - 8.1.1. Master Data Definition
  - 8.1.2. Sources of Master Data
  - 8.1.3. Types of Master Data
  - 8.1.4. MDM Objectives
  - 8.1.5. MDM Architectures and Styles
  - 8.1.6. Costs and Benefits of MDM
  - 8.1.7. MDM Challenges and Best Practices
- 8.2. *Data Parsing, Matching, and De-Duplication*
  - 8.2.1. Data Parsing and Standardization
    - 8.2.1.1. Personal and Business Name
    - 8.2.1.2. Postal Addresses
    - 8.2.1.3. Geocoding

- 8.2.1.4. Date and Time
- 8.2.2. Data Matching and De-Duplication
  - 8.2.2.1. Deterministic vs. Probabilistic
  - 8.2.2.2. Matching Rules
  - 8.2.2.3. Data Survivorship
  - 8.2.2.4. Types of Matching Tools
- 8.3. *Reference Data*
  - 8.3.1. External Reference Data
    - 8.3.1.1. Address Directories
    - 8.3.1.2. Business Directories
    - 8.3.1.3. DUNS and EPC
  - 8.3.2. Internal Reference Data
    - 8.3.2.1. Code and Lookup Tables
    - 8.3.2.2. Valid Values Lists
    - 8.3.2.3. Demographic Fields
    - 8.3.2.4. Standard Abbreviations
    - 8.3.2.5. Language Lists
    - 8.3.2.6. Currencies
- 8.4. *Global Data*
  - 8.4.1. Format Differences across Countries and Regions
    - 8.4.1.1. Address Formats
    - 8.4.1.2. Personal Names
    - 8.4.1.3. Business Names
    - 8.4.1.4. Phone Numbers
    - 8.4.1.5. Job Titles
  - 8.4.2. Special Characters and Diacritics
    - 8.4.2.1. Regional Alphabets
    - 8.4.2.2. Character Sets
    - 8.4.2.3. Code Pages and Unicode
  - 8.4.3. Cultural and Legal Issues
    - 8.4.3.1. Privacy Laws
    - 8.4.3.2. Cultural Differences
  - 8.4.4. Best Practices of Working with Global Data

**Note:** While this document goes into four levels of detail, the fourth level is not exhaustive and in many cases only includes some examples for topic clarification.